

# A Computational Framework for Systems-based Analysis of Developmental Toxicity

Thomas B. Knudsen(1), Amar V. Singh(2), Robert J. Kavlock(1), David J. Dix(1), Richard S. Judson(1), Matthew Martin(1), Vicki Dellarco(3), Elizabeth Mendez(3) and Elissa Reaves (3)  
(1)NCCT/ORD, USEPA, RTP NC, USA, (2) Lockheed Martin, RTP, USA, (3)OPP/OPPTS, USEPA, Washington DC, USA

## 1. INTRODUCTION

**Motivation:** Developing systems are often sensitive to chemicals below levels that adversely affect the pregnant mother. Prenatal studies are among the largest consumer of animal resources in toxicity testing and the number of chemicals requiring testing far exceeds what has and can be delivered through traditional methods [1].

**v-Embryo:** EPA recently launched 'The Virtual Embryo Project' (v-Embryo) as a research program to develop a sophisticated computer model of a human embryo that will ultimately help to understand and predict developmental risks posed by environmental chemicals to the embryo, fetus or newborn [2]. It's two major computational modules focus on 'knowledge discovery' and 'systems modeling'.

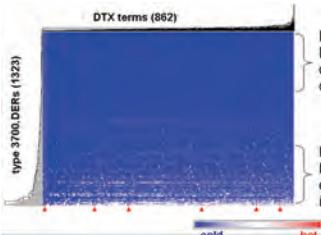
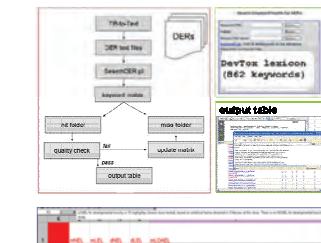
**Objective:** The work reported here establishes a flexible and extensible infrastructure for knowledge-discovery that draws from empirical observations and inference algorithms for reasoning about developmental toxicity (DTX).

## 2. WORKFLOW

**Documents:** EPA's ToxRefDB database makes animal toxicity data accessible for correlation with bioactivity profiling data from the ToxCast™ high-throughput assays [3]. ToxRefDB structures information from Data Evaluation Records (DERs), which are reviews of registrant-submitted toxicity studies at EPA's Office of Pesticide Programs, into computable form.

**Document library:** A searchable repository was derived from 4618 DERs used for ToxRefDB to date. We converted 3775 imaged \*.tif files using IdentifyDocs software and 837 \*.wp files using ABC Amber WordPerfect Converter software, with ~85-90% fidelity.

**Text-mining:** To assist manual curation of structured data and to extract additional detail, concepts and ideas from the document library, we wrote a shallow parser in Perl to search by features relevant to DTX.



## 4. ANALYSIS

Of 1323 prenatal DERs (391 chemicals), 981 (74%) hit at least 1 DTX feature across 376 (96%) chemicals. Feature representation (above) is displayed by region and phenotype (top 30 and bottom 30).

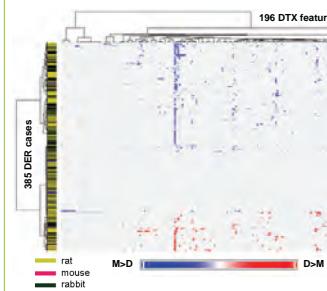
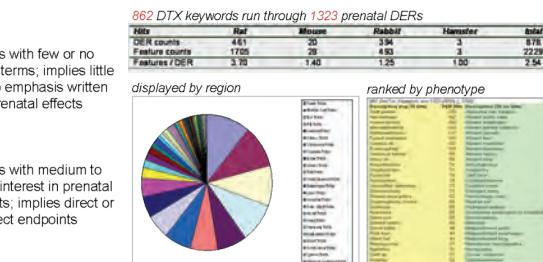
For 385 cases, M/D score exceeded log2 = +0.75 in 50 cases (13.0%) and log2 = -0.75 in 29 cases (7.5%); these represent studies provisionally classified as high and low potential for developmental relative to maternal toxicities. For feature mapping (right), we clustered by M/D score over the 385 DER cases for 196 DTX features.

## 3. OUTPUT

The shallow parser screened the DER library using a lexicon of 862 DTX terms (<http://www.devtox.org>). A string of semi-automated steps fetched, parsed, and formatted the unstructured data from 1323 prenatal studies.

The number of occurrences of each keyword in every file was automatically written to a summary file. The interactive table returns text and original DERs for manual inspection. A sample is shown for maternal and developmental NELs and LOELs.

Information on lowest effective dose (LEL, LOEL, LOAEL) and no effect level (NEL, NOEL, NOAEL) was parsed for determining the Maternal / Developmental values, and a log2 M/D score was computed for 385 cases having both.



## 5. SUMMARY

Automated text-mining can help analyze informational patterns for document summarization and clustering.

Tools to capture major themes or ideas from literature and legacy records will enhance a DTX knowledgebase.

Data-mining and text-mining techniques together can and assist, accelerate, or QC manual data entry (ToxRefDB).

Full implementation of automated text-mining will require a systems-based DevTox ontology and knowledgebase.

v-Embryo aims to integrate knowledge-discovery tools with experimental data and systems modeling efforts.

## 6. REFERENCES

1. NRC Committee on Toxicity and Assessment of Environmental Agents. Toxicity Testing in the Twenty-first Century: A Vision and a Strategy (2007). National Academies Press (<http://www.nap.edu/catalog/11970.html>).
2. Risk Policy Report (2008). EPA high-tech 'Virtual Embryo Project' will target developmental risk. vol 15, no. 2, January 8, 2008. (<http://www.InsideEPA.com>).
3. Dix DJ, Houck KA, Martin MT, Richard AM, Setzer RW and Kavlock RJ (2007) The ToxCast program for prioritizing toxicity testing of environmental chemicals. *Toxicol Sci* 95: 5-12.

*Disclaimer: This work has been reviewed by EPA and approved for presentation but does not necessarily reflect the views and policies of the Agency.*